

# Optimal Rate Allocation in Mismatched Multiterminal Source Coding

Ruiyang Song<sup>\*</sup>, Stefano Rini<sup>†</sup>, Alon Kipnis<sup>‡</sup>, and Andrea J. Goldsmith<sup>‡</sup>

<sup>\*</sup> Tsinghua University, Beijing, China

<sup>†</sup> National Chiao-Tung University, Hsinchu, Taiwan

<sup>‡</sup> Stanford University, CA, USA

Emails: songry12@mails.tsinghua.edu.cn, stefano@nctu.edu.tw, kipnisal, andrea@wsl.stanford.edu

## Abstract

We consider a multiterminal source coding problem in which a source is estimated at a central processing unit from lossy-compressed remote observations. Each lossy-encoded observation is produced by a remote sensor which obtains a noisy version of the source and compresses this observation minimizing a local distortion measure which depends only on the marginal distribution of its observation. The central node, on the other hand, has knowledge of the joint distribution of the source and all the observations and produces the source estimate which minimizes a different distortion measure between the source and its reconstruction. In this correspondence, we investigate the problem of optimally choosing the rate of each lossy-compressed remote estimate so as to minimize the distortion at the central processing unit subject to a bound on the overall communication rate between the remote sensors and the central unit. We focus, in particular, on two models of practical relevance: the case of a Gaussian source observed in additive Gaussian noise and reconstructed under quadratic distortion, and the case of a binary source observed in bit-flipping noise and reconstructed under Hamming distortion. In both scenarios we show that there exist regimes under which having more remote encoders does reduce the source distortion: in other words, having fewer, high-quality remote estimates provides a smaller distortion than having more, lower-quality estimates.

## Index Terms

Remote source coding; Binary source; Gaussian source; Binary symmetric channel; CEO problem.

## I. INTRODUCTION

The Distortion-Rate Function (DRF) characterizes the minimum code-rate required for encoding an information source as to ensure its recovery to be within a prescribed average distortion. A natural extension of this setup is the case in which the encoder can only observe the source through noisy observations: this scenario is referred to as the *remote* or *indirect* source coding problem [1, Sec. 3.5]. The setting in which multiple remote encoders aid the reconstruction of a remote source at a central decoder corresponds to the Chief Executive Officer (CEO) problem [2]. Unfortunately, the CEO problem is intrinsically arduous to solve. Hence, despite its relevance in many sensing applications, only a few results for this model are available in the literature [3], [4].

Given the difficulty in deriving the optimal rate-distortion trade-off in the CEO problem, one naturally wonders if there exists a more tractable model which addresses the multiterminal remote source coding problem. Indeed, the mismatched multiterminal source coding problem in [5] is a step in this direction as it considers the case in which the distortion criterion at the remote encoders does not depend on the source distribution or the decoder distortion measure. This is a different setup than the CEO problem, but a judicious choice of the distortion measure for the remote encoders produces similar performances [5]. Since the codebooks employed by the remote encoders are not necessarily optimal in minimizing the distortion criterion at the central node, this problem is referred to as the *mismatched* multiterminal source coding. In the mismatched multiterminal source coding problem, the encoding operations at the remote encoders are chosen among a family of possible encoding rules and the reconstruction operation is determined by the optimal source estimate given the remote lossy-compressed samples. Specifically, it is assumed that these encoding rules are chosen to operate at the optimal rate-distortion point asymptotically as the block length goes to infinity with respect to the local distortion measure at each encoder. Using properties of distributions that achieve the DRF, it was shown in [5] that the minimal distortion in reconstructing the source approaches a single letter expression as the blocklength of the encoders goes to infinity. The optimal distortion-rate trade-off for this scenario is described by the mismatched Distortion Rate Function (mDRF). The availability of such a single letter expression for the mDRF allows us to evaluate the distortion-rate trade-off in closed form and moreover, as explored in this paper, to determine the optimal rate-allocation strategy given a prescribed target distortion.

The optimal rate-allocation in the mismatched multiterminal problem is of great practical relevance in distributed sensing networks, in which the remote sensors are either unaware of the existence of the underlying source or lack the flexibility and the resources to adapt to multiple sensing scenarios. Under this setting, the sensors may have very different quality of

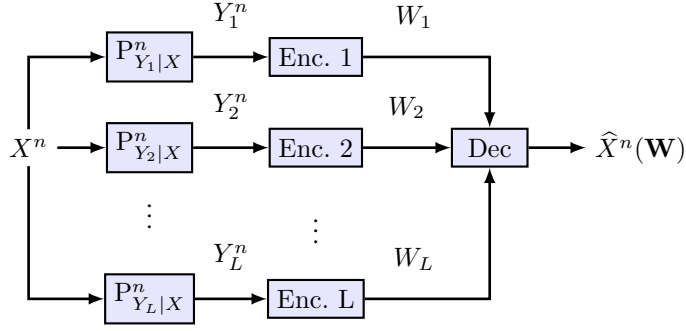


Fig. 1: The multiterminal mismatched distortion rate problem.

measurements, forcing the central processor to allocate different communication resources and associated data rates to different sensors.

The remainder of this paper is organized as follows: in Sec II we introduce the problem formulation, while in Sec. III we present relevant results available in the literature. The result for the Gaussian source is analyzed in Sec. IV; the case for the binary source is analyzed in Sec. V. Finally, Sec. VI concludes the paper.

## II. PROBLEM FORMULATION

Our starting point is the source coding setting of Fig. 1: the source sequence  $X^n = (X_1, \dots, X_n)$  is obtained through  $n$  iid draws from the distribution  $P_X(x)$ . This source is observed at  $L$  remote encoders through  $n$  uses of the memoryless channel  $P_{Y_l|X}$ , ie:

$$P_{Y_l^n|X^n}(y_l^n|x^n) = \prod_{i=1}^n P_{Y_l|X}(y_{l,i}|x_i), \quad l \in \{1, \dots, L\}. \quad (1)$$

Given the observation  $Y_l^{(n)}$ , each remote encoder  $\mathcal{E}_l^{(n)}$  produces an index  $W_l \in \{1 \dots 2^{\lfloor nR_l \rfloor}\}$ . The decoder receives  $\mathbf{W} = (W_1, \dots, W_L)$  and produces the sequence  $\hat{X}^n(\mathbf{W})$ . We consider the distortion between the reconstruction  $\hat{X}^n$  and the original source realization defined as

$$D^n \leq D(X^n, \hat{X}^n) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{E} [d(X_i, \hat{X}_i)], \quad (2)$$

for some positive defined per-letter distortion  $d(x, \hat{x})$ . Given  $\mathbf{W}$ , the decoder produces a reconstruction sequence  $\hat{X}^n$  minimizing (2).

An achievable distortion and the optimal distortion-rate trade-off are defined as follow: for a family of encoders of block-length  $n$ :  $\mathcal{E}^{(n)} \triangleq \mathcal{E}_1^{(n)} \times \dots \times \mathcal{E}_L^{(n)}$ , and an  $n$ -sequence distortion measure  $D^n$ , we introduce the following definition:

**Definition 1.** A distortion  $D$  is achievable for encoders  $\mathcal{E}^{(n)}$  if there exist  $L$  encoders  $(g_1, \dots, g_L) \in \mathcal{E}^{(n)}$  such that  $D^n \leq D$ .

The mDRF arises when, for block-length  $n$ , the encoders  $\mathcal{E}^{(n)}$  are chosen as in the classical source coding problem with source  $Y_l^n$  and distortion criterion

$$D_l^n(Y_l^n, \hat{Y}_l^n) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{E} [d_l(y_i, \hat{y}_i)], \quad (3)$$

for some per-letter distortion  $d_l(y, \hat{y})$ . More precisely, we require that the family of encoders  $\mathcal{E}^{(n)}$  is such that the optimal distortion-rate performance with respect to the sequence  $Y_l^n$  and distortion  $d_l$  is asymptotically achieved for all  $l = 1, \dots, L$ . Namely,  $D_l^n(Y_l, \hat{Y}_l) \rightarrow D_l(R_l)$  while  $I(Y_l; \hat{Y}_l) \rightarrow R_l$  where  $D_l(R_l)$  is the DRF for the problem of reconstructing  $Y$  under per-letter distortion  $d_l$  at rate  $R_l$ . Under these conditions, it is possible to obtain the following characterization of the distortion in estimating the source  $X^n$  under this family of encoders, denoted by  $D(\mathcal{E}(R_1 \dots R_L))$  [5, Th. 4.1]:

$$D(\mathcal{E}(R_1 \dots R_L)) \triangleq \inf D(X, \hat{X}), \quad (4)$$

where the infimum is over all possible random mappings  $(\hat{Y}_1 \dots \hat{Y}_L) \rightarrow \hat{X}$ , where  $p(\hat{y}_1|y_1), \dots, p(\hat{y}_L|y_L)$  are  $L$  conditional distributions for which  $I(Y_l; \hat{Y}_l) = R_l$ , and  $D_l(Y_l, \hat{Y}_l) = D_l(R_l)$ . We will refer to the function  $D(\mathcal{E}(R_1 \dots R_L))$  as the *mismatched multiterminal* distortion-rate function (mDRF).

Note that, unlike the traditional source coding definitions, the mDRF is not defined as an optimization of encoding-decoding schemes with respect to the source  $X$ . It is only defined as the attainable minimal distortion once the  $L$ ,  $n$ -block-length

encoders are determined. The only requirement for these encoders is that, as  $n$  goes to infinity, they achieve the rate-distortion function with respect to their local distortion measure and input sequence. The problem considered in this correspondence is as follows: given a total communication rate-budget of  $R$  bits, we ask what is the minimal value of  $D(\mathcal{E}(R_1 \dots R_L))$  subject to the constraint

$$\sum_{l=1}^L R_l \leq R, \quad (5)$$

for some  $R \in \mathbb{R}^+$ .

We focus, in particular, on two models:

- **Gaussian source in Gaussian noise and quadratic distortion:** the source  $X \sim \mathcal{N}(0, 1)$  and the  $l$ th observation is

$$Y_l = \sqrt{\gamma_l} X + Z_l, \quad l \in [1, \dots, L], \quad (6)$$

for  $Z_l \sim \mathcal{N}(0, 1)$  and iid with  $0 \leq \gamma_L \leq \dots \leq \gamma_1$ , while  $d_l(u, \hat{u}) = d(u, \hat{u}) = (u - \hat{u})^2$ .

- **Binary source in bit-flipping noise and Hamming distortion:** the iid source follows  $\text{Bern}(1/2)$  and the  $l$ th observation is

$$Y_l = X \oplus Z_l, \quad l \in [1, \dots, L], \quad (7)$$

for  $Z_l \sim \text{Bern}(p_l)$  with  $0 \leq p_1 \leq \dots \leq p_L$  while  $d_l(u, \hat{u}) = d(u, \hat{u}) = u \oplus \hat{u}$ .

**Notations:** In the following sections, we denote  $\bar{x} = 1 - x$ , and  $x \star y = x\bar{y} + \bar{x}y$ . Moreover  $h$  denotes the binary entropy, i.e.  $h(x) = x \log x + \bar{x} \log \bar{x}$ . With  $h^{-1}(x)$  for  $0 \leq x \leq 1/2$  we indicate the inverse of the binary entropy function, which is well defined in this interval.

### III. RELATED RESULTS

The source coding setting in which a single encoder only has partial information about the source is usually studied in the setting of universal source coding [6, Ch. 11.3]. Another approach is to encode with respect to a min-max penalty over a family of source distributions [7]. Mismatched encoding can also arise in the case in which the codebook at the decoder is fixed but its encoding codebook is not [8].

The most related multiterminal setting to this work is the CEO problem introduced in [2]. The CEO setup corresponds to the case where for each  $n$ , the optimization over (2) at the  $l$ th encoder is with respect to the entire family of block-length  $n$  encoders  $\mathcal{E}_l^{(n)}$ . The relevant results for the CEO in the quadratic Gaussian case and the binary symmetric case are as follow:

- **quadratic Gaussian CEO problem.** The rate-region for the Gaussian CEO setup was derived in [3], while the optimal rate-allocation is studied in [9]. For the case of  $\gamma = \gamma_1 = \dots = \gamma_L$ , the optimal rate-allocation for a given sum rate constraint  $\sum R_l \leq R$  is given by [9, Eq. 10]:

$$R(D^*) = \frac{1}{2} \log^+ \left( \frac{1}{D^*} \right) - \frac{L}{2} \log^+ \left( 1 + \frac{1}{\alpha^2} - \frac{1}{\alpha^2 D^*} \right). \quad (8)$$

Note that the difference in our setup from that of the CEO problem of [2] is that, in our problem formulation, the  $l$ th encoder is limited to a specific family  $\mathcal{E}_l^{(n)}$  of block-length  $n$  encoders at rate  $R_l$  which have optimal performance with respect to a local distortion measure in terms of the observable process  $Y_l^n$ .

- **Hamming binary multiterminal mismatched problem.** The channel model in (7) is studied in [10] where an lower bound for the rate-distortion trade-off for the case of  $L = 2$  is obtained as

$$R_1 \geq h(\rho \star h^{-1}(1 - R_2)) - h(D_1) \quad (9a)$$

$$R_2 \geq h(\rho \star h^{-1}(1 - R_1)) - h(D_2) \quad (9b)$$

$$R_1 + R_2 \geq 1 + h(\rho) - h(D_1) - h(D_2) \quad (9c)$$

for  $\rho = p_1 \star p_2$ , and  $R_1, R_2 \leq 1$ .

### IV. GAUSSIAN SOURCE WITH GAUSSIAN NOISE AND QUADRATIC DISTORTION

As a first example of the optimal rate allocation in the mismatched multiterminal source coding problem, we consider the model in (6) for which the mDRF is shown in [5, Prop. 5.2].

**Proposition IV.1.** *The mDRF in the distributed encoding of a Gaussian source with observation model formulated in (6) is given by*

$$D(\mathcal{E}(R_1 \dots R_L)) = \left( 1 + \sum_{l=1}^L \gamma_l \frac{1 - 2^{-2R_l}}{1 + \gamma_l 2^{-2R_l}} \right)^{-1}. \quad (10)$$

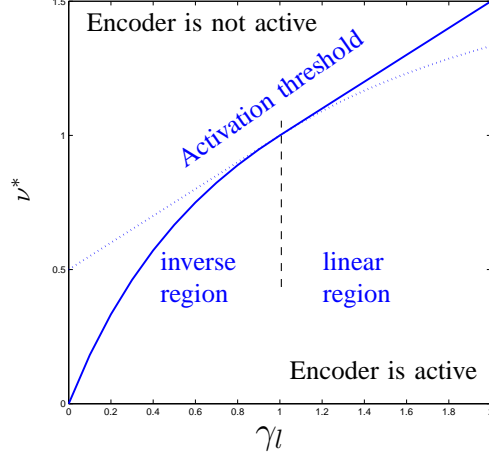


Fig. 2: The remote user activation threshold in Th. IV.2.

With Prop. IV.1, the optimal rate allocation under the sum rate constraint of (5) is obtained in the next theorem.

**Theorem IV.2.** Let  $L_0$  be the number of coefficients  $\gamma_l$  equal to  $\gamma_1$ , that is  $L_0 = \sum_{i=1}^L 1_{\{\gamma_i = \gamma_1\}}$ , and define

$$g(\nu, \gamma_l) = \frac{\gamma_l}{\nu} \left( \gamma_l + 1 - \nu + \sqrt{(\gamma_l + 1)(\gamma_l + 1 - 2\nu)} \right). \quad (11)$$

When

$$R \geq \frac{L_0}{2} \log^+ \gamma_1, \quad (12)$$

and if there exists a  $\nu^*$  such that

$$\sum_{l=1}^L R_l^*(\nu^*) = R, \quad (13)$$

then, the optimal rate allocation for the model in (6) under the sum rate constraint in (5) is attained by the following choice:

$$R_l^*(\nu) = \begin{cases} \frac{1}{2} \log g(\nu^*, \gamma_l) & 0 \leq \gamma_l \leq 1, \quad 0 \leq \nu^* \leq \frac{2\gamma_l}{\gamma_l + 1} \\ \frac{1}{2} \log g(\nu^*, \gamma_l) & 1 \leq \gamma_l, \quad 0 \leq \nu^* \leq \frac{\gamma_l + 1}{2} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

*Proof:* The proof is provided in App. A. ■

Note that (12) is always satisfied if  $\gamma_1 \leq 1$ . However, for  $\gamma_1 > 1$ , the solution  $\nu^*$  may not exist for small  $R$ . Moreover, we see that if  $\gamma_1 > \gamma_i > 1$  for some  $i$ , the sum rate constraint (13) may still not have a solution for some values of  $R$  due to the discontinuity of  $\sum_{l=1}^L R_l^*(\nu^*)$ . The result in Th. IV.2 can be interpreted as follows: we refer to a remote user which transmits at a rate strictly larger than zero as “active”. Then the conditions in the r.h.s. of (14) decide which encoders are active as a function of the parameter  $\nu^*$ . For a given  $\nu^*$ , while (14) decides whether a user is active or not, (11) specifies the rates of the active users. The parameter  $\nu^*$  is chosen so that the total rate constraint is met with equality when such a solution exists.

An interesting observation is that in (14) a simple threshold decides which encoders are active: for the remote encoders with  $0 \leq \gamma_l \leq 1$ , the activation rule is  $\nu^* \leq 2\gamma_l/(\gamma_l + 1)$  (inverse in  $\gamma_l$ ); while for encoders with  $\gamma_l > 1$ , we have the activation rule as  $\nu^* \leq (\gamma_l + 1)/2$  (linear in  $\gamma_l$ ). The activation rules are illustrated in Fig. 2.

The optimal rate allocations for the case of  $L = 5$  users with various rate budgets are presented in Fig. 3. Note that the low-SNR remote encoders are active only for very large rate budgets and are quickly deactivated as the total rate budget decreases. The result in Fig. 3 is obtained by noticing that (11) is decreasing in  $\nu$ , hence the optimal allocation is numerically obtained by setting  $\nu = (\gamma_1 + 1)/2$  and progressively reducing this value until the sum rate constraint in (5) is met with equality.

Another scenario we are interested in is the case when the optimal rate allocation only has one active encoder and thus the presence of the second encoder does not further decrease the distortion. For  $L = 2$ , recall that (10) is a function of  $R_1$ . By minimizing (10), we can then derive an upper bound for the total rate budget below which only one sensor should be active in order to optimally represent the hidden source. In Fig. 4 we plot this value of  $R$  as a function of  $\gamma_2$  for various  $\gamma_1$ . If the total budget is below the plotted line, then the optimal rate allocation is  $R_1^* = R$ ,  $R_2^* = 0$ .

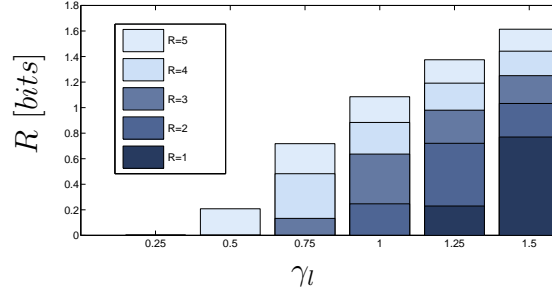


Fig. 3: The optimal rate allocation for the case of  $L = 5$  and for different values of  $\gamma_l$  and  $R$ .

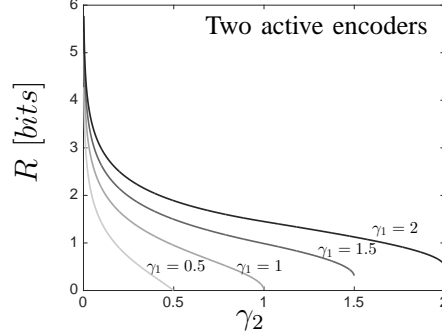


Fig. 4: The total rate budget below which all the rate should be allocated to one user for the case of Gaussian source,  $L = 2$ , and  $\gamma_1 \in \{.5, 1, 1.5, 2\}$ ,  $\gamma_2 \in [0, \gamma_1]$ .

## V. BINARY SOURCE WITH BIT-FLIPPING NOISE UNDER HAMMING DISTORTION

Another model of interest for many practical communication settings is the one in which a binary source is observed through Binary Symmetric Channels (BSCs) at the remote sensors and reconstructed at the central unit by minimizing the Hamming distortion. In order to derive the optimal rate-allocation for the model formulated in (7), we first write the mDRF derived in [5, Prop. 5.3] for the case  $\mathbb{P}(X_i = 1) = 1/2$  in the following form:

**Proposition V.1.** *The mDRF in the distributed encoding of a binary source with the observation model specified in (7) is*

$$D(\mathcal{E}(R_1 \dots R_L)) = \mathbb{P}(\log F > 0), \quad (15)$$

where

$$\log F = \sum_{l=1}^L c_l (2U_l - 1), \quad (16)$$

for  $c_l = \log \frac{1-q_l}{q_l}$  and where  $U_1, \dots, U_L$  are i.i.d binary random variables satisfying

$$\mathbb{P}(U_l = 1) = q_l = p_l \star D_l, \quad (17)$$

where  $D_l = h^{-1}(1 - R_l)$ .

*Proof:* The proof is a reformulation of the result in [5, Prop. 3]. ■

Proposition V.1 provides the following interpretation for the mDRF: each lossy-compressed observation  $\hat{Y}_l$  is obtained by passing the observation  $X$  through a series of two BSCs: one with cross-over probability  $p_l$  and the other with cross-over probability  $D_l$ , corresponding to a BSC with cross-over probability  $q_l$  as defined in (17). The BSC with cross-over probability  $D_l$  corresponds to the optimal test-channel in lossy compression of a binary uniform source under Hamming distortion at rate  $R_l(D_l)$ . The optimal estimation of  $X$  at the central decoder is obtained through a log-likelihood ratio test in which each observation is weighted according to a parameter  $c_l$  which is controlled by the cross-over probability  $q_l$ . Thus, the distortion corresponds to the probability that the log-likelihood ratio fails given the two equally likely observations.

The following theorem describes the optimal rate allocation that minimizes (15):

**Theorem V.2.** *The optimal rate allocation for the model in (7) under the sum rate constraint in (5) is attained by the following*

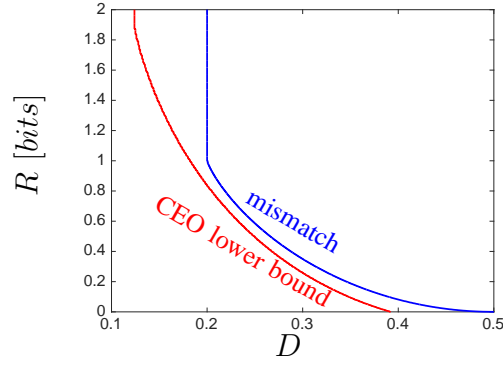


Fig. 5: The optimal rate allocation for symmetric binary CEO lower bound (red) and the symmetric mismatched remote source coding problem (blue) for  $L = 2$  and  $p = 0.2$  in Sec. V.

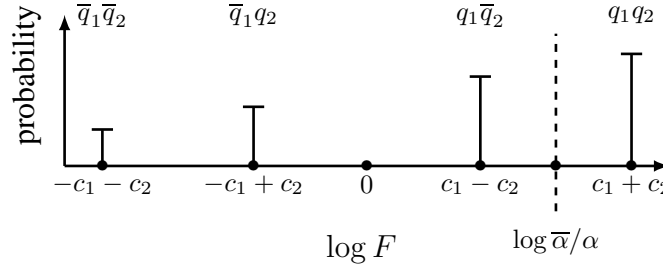


Fig. 6: An example of  $\log F$  for the case in which the source is Bernoulli with parameter  $\alpha$ .

choice:

$$R_l^* = \begin{cases} 0 & R < l - 1 \\ R - l + 1 & l - 1 \leq R < l \\ 1 & R \geq l \end{cases} \quad (18)$$

*Proof:* The result in Th. V.2 is shown by considering the necessary properties of the optimal solution. In particular, if a solution is optimal, then there cannot be a remote encoder observing the source through a BSC with cross-over probability  $p_{l+1}$  with a positive rate allocated while there also exists an encoder observing the source through a BSC with  $p_l < p_{l+1}$  and  $R_l < 1$ . This is because a lower overall distortion would be attained by assigning the rate of user  $l + 1$  to user  $l$  until  $R_l = 1$ . The remainder of the proof can be found in App. B. ■

As in the result of Th. IV.2, the result in Th. V.2 shows that there exists regimes in which a remote encoder is ignored by the optimal rate allocation if its observation is too noisy or if the available communication rate is too small.

Note that the rate for the symmetric case  $p_1 = \dots = p_l = p$  does not result in the same rate being assigned to all the remote encoders. This is in stark contrast with the optimal solution for the CEO problem: in Fig. 5 we illustrate the symmetric binary case of (18) and compare it with the CEO result in (9c).

Th. V.2 holds for a symmetric source: if the source follows a general Bernoulli distribution with parameter  $\alpha < 1/2$ , then the solution in Th. V.2 is no longer optimal. This can be clearly shown for the case  $L = 2$ : in this case the distribution of  $\log F$  is represented in Fig. 6. We can see that the support of  $\log F$  has 4 points:  $(-c_1 - c_2, -c_1 + c_2, +c_1 - c_2, +c_1 + c_2)$ , with probability  $(\bar{q}_1\bar{q}_2, \bar{q}_1q_2, q_1\bar{q}_2, q_1q_2)$  respectively.

Assuming that  $c_1 > c_2$ , the log-likelihood threshold  $\log \bar{\alpha}/\alpha$  corresponds to a positive value which can be located either in the interval  $[0, +c_1 - c_2]$  or in  $[+c_1 - c_2, c_1 + c_2]$ . In the first interval, the probability of error is  $q_1$ , regardless of the value of  $q_2$  while, in the second interval, the probability of error is  $q_1q_2$ . When minimizing the probability  $q_1$ , the best rate allocation strategy is to assign the full rate to  $R_1$  while the rate assigned to  $R_2$  does not affect the overall distortion. In general, this choice is not optimal when minimizing  $q_1q_2$ . Since determining the optimal choice for the latter case in closed form is challenging, we show this result numerically in Fig. 7.

## VI. CONCLUSIONS

We have considered the optimal rate allocation schemes for two mismatch multiterminal source coding problems. In particular, for an iid Gaussian source observed through AWGN channels under a quadratic distortion measure, we derived the optimal allocation scheme in which remote users are assigned a positive rate according to a threshold that varies with the observation

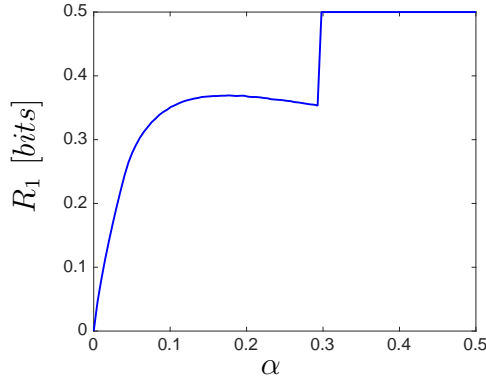


Fig. 7: The optimal rate allocation of  $R_1$  for a Bernoulli source with parameter  $\alpha \in [0, 1/2]$  and the total sum-rate  $R = 0.5$  while the cross-over probabilities are  $p_1 = 1/5$  and  $p_2 = 1/3$ .

quality. For an iid symmetric binary source observed through binary symmetric channels under Hamming distortion, we also determine the optimal rate allocation and, for this model, the remote encoders with less noisy observations compress their observations at the minimum distortion and thus the number of active encoders equals the number of total bits plus one.

#### REFERENCES

- [1] T. Berger, *Rate-distortion theory: A mathematical basis for data compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
- [2] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem [multiterminal source coding]," *Information Theory, IEEE Transactions on*, vol. 42, no. 3, pp. 887–902, 1996.
- [3] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic gaussian ceo problem," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*. IEEE, 2004, p. 119.
- [4] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *IEEE Transactions on Information Theory*, vol. 60, no. 1, pp. 740–761, Jan 2014.
- [5] A. Kipnis, S. Rini, and A. J. Goldsmith, "Mismatched multiterminal source coding," 2016, submitted to ISIT 2016, Barcelona. [Online]. Available: <http://arxiv.org/abs/1602.02201>
- [6] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
- [7] A. Dembo and T. Weissman, "The minimax distortion redundancy in noisy source coding," *Information Theory, IEEE Transactions on*, vol. 49, no. 11, pp. 3020–3030, 2003.
- [8] A. Lapidoth, "On the role of mismatch in rate distortion theory," *Information Theory, IEEE Transactions on*, vol. 43, no. 1, pp. 38–47, 1997.
- [9] J. Chen, X. Zhang, T. Berger, and S. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *Selected Areas in Communications, IEEE Journal on*, vol. 22, no. 6, pp. 977–987, Aug 2004.
- [10] X. He, X. Zhou, P. Komulainen, M. Juntti, and T. Matsumoto, "A lower bound analysis of hamming distortion for a binary CEO problem with joint source-channel coding," 2015.

## APPENDIX

### A. Proof of Theorem IV.2

Let  $\mathbf{R} = (R_1, \dots, R_L)$ . The optimal allocation of rate  $\mathbf{R}^*$  minimizes the minimum mean square error in (10) and hence is the solution to the optimization problem which can be formulated as follows:

$$\begin{aligned} f_L(\mathbf{R}) &\triangleq - \left( 1 + \sum_{l=1}^L \frac{\gamma_l(2^{2R_l} - 1)}{\gamma_l + 2^{2R_l}} \right), \\ \text{subject to } & \mathbf{1}^\top \mathbf{R} \leq R, \\ & \mathbf{R} \succeq 0. \end{aligned} \quad (19)$$

The first and second order partial derivatives of the target function  $f_L(\mathbf{R})$  can be expanded as

$$\frac{\partial f_L}{\partial R_l} = -2\gamma_l(\gamma_l + 1) \frac{4^{R_l}}{(4^{R_l} + \gamma_l)^2}, \quad (20)$$

$$\frac{\partial^2 f_L}{\partial R_l^2} = 4\gamma_l(\gamma_l + 1) \frac{4^{R_l}(4^{R_l} - \gamma_l)}{(4^{R_l} + \gamma_l)^3}, \quad (21)$$

and

$$\frac{\partial^2 f_L}{\partial R_l \partial R_k} = 0, \text{ for } l \neq k. \quad (22)$$

The Lagrangian of the optimization problem in (19) is

$$L(\mathbf{R}, \boldsymbol{\lambda}, \nu) = f_L(\mathbf{R}) + \sum_{l=1}^L \lambda_l(-R_l) + \nu(\sum_{l=1}^L R_l - R). \quad (23)$$

Let  $\mathbf{R}^*$  and  $(\boldsymbol{\lambda}^*, \nu^*)$  be the primal and dual optimal points with zero duality gap. The KKT conditions can be expanded as follow:

$$\mathbf{R}^* \succeq 0, \quad (24a)$$

$$\sum_{l=1}^L R_l^* - R = 0, \quad (24b)$$

$$\boldsymbol{\lambda}^* \succeq 0, \quad (24c)$$

$$\nu^* \geq 0, \quad (24d)$$

$$\lambda_l^* R_l^* = 0, \quad l = 1, \dots, L, \quad (24e)$$

$$\frac{\partial f_L(R^*)}{\partial R_l} - \lambda_l^* + \nu^* = 0, \quad l = 1, \dots, L. \quad (24f)$$

When a rate  $R_l > 0$ , we must have  $\lambda_l = 0$  because of (24e), in which case (24f) yields

$$\nu^* - 2\gamma_l(\gamma_l + 1) \frac{4^{R_l^*}}{(4^{R_l^*} + \gamma_l)^2} = 0. \quad (25)$$

The solution of (25) in  $4^{R_l^*}$ , if it exists, is

$$4^{R_l^*} = \frac{\gamma_l}{\nu^*} \left( \gamma_l + 1 - \nu^* \pm \sqrt{(\gamma_l + 1)(\gamma_l + 1 - 2\nu^*)} \right) \triangleq g_{\pm}(\nu^*, \gamma_l) \quad (26)$$

for

$$0 \leq \nu^* \leq \frac{\gamma_l + 1}{2}. \quad (27)$$

Note that

$$\begin{aligned} \frac{\partial^2 f_L}{\partial R_l^2} \left( \frac{1}{2} \log(g_+(\nu^*, \gamma_l)) \right) &> 0, \\ \frac{\partial^2 f_L}{\partial R_l^2} \left( \frac{1}{2} \log(g_-(\nu^*, \gamma_l)) \right) &< 0, \end{aligned}$$

so that only the  $g_+$  solution corresponds to the minimal of the target function. We can therefore drop the subscript  $+$  in (26) and simply refer to  $g_+(\nu^*, \gamma_l)$  as  $g(\nu^*, \gamma_l)$  as in (11).



The solution in (26) corresponds to a positive rate only when  $g(\nu^*, \gamma_l) \geq 1$ , which can be equivalently written as

$$\gamma_l \sqrt{\gamma_l + 1 - 2\nu^*} \geq (\nu^* - \gamma_l) \sqrt{\gamma_l + 1}. \quad (28)$$

This is always the case when  $\gamma_l > 1$ . However, when  $\gamma_l < 1$  it requires the additional condition

$$0 \leq \nu \leq \frac{2\gamma_l}{\gamma_l + 1}. \quad (29)$$

Since

$$\frac{\gamma_l + 1}{2} \geq \frac{2\gamma_l}{\gamma_l + 1}, \quad (30)$$

we have that the optimal rate allocation is

$$R_l^*(\nu) = \begin{cases} \frac{1}{2} \log g(\nu^*, \gamma_l) & 0 \leq \gamma_l \leq 1 & 0 \leq \nu^* \leq \frac{2\gamma_l}{\gamma_l + 1} \\ \frac{1}{2} \log g(\nu^*, \gamma_l) & \gamma_l \geq 1 & 0 \leq \nu^* \leq \frac{\gamma_l + 1}{2} \\ 0 & & \text{otherwise} \end{cases} \quad (31)$$

where  $\nu^*$  is determined by the following equation:

$$\sum_{l=1}^L R_l^*(\nu) = R_T(\nu) = R. \quad (32)$$

Additionally, since  $\frac{\partial g(\nu, \gamma_l)}{\partial \nu} < 0$ , we also have that

$$g(\nu, \gamma_l) \geq g\left(\frac{\gamma_l + 1}{2}, \gamma_l\right) = \gamma_l. \quad (33)$$

Note that (32) can be satisfied only for certain values of  $R$ . More specifically, (33) and (31) together imply

$$R_l \geq \frac{1}{2} \log(\gamma_l), \quad (34)$$

which corresponds to the conditions under which the problem is convex. The bound in (34) is not tight for  $\gamma_l < 1$  but could possibly be tight for  $\gamma_l > 1$ . Fortunately, we have

$$g(\nu, \gamma_l) \geq \gamma_l, \quad \nu \in \left[0, \frac{\gamma_l + 1}{2}\right] \quad (35)$$

since

$$g(\nu, \gamma_l) - \gamma_l = \frac{\gamma_l}{\nu} \left( \gamma_l + 1 - 2\nu + \sqrt{(\gamma_l + 1)(\gamma_l + 1 - 2\nu)} \right), \quad (36)$$

which is positive defined for  $\nu \in [0, \frac{\gamma_l + 1}{2}]$ . Therefore, we will need the sum rate to be lower bounded by  $\frac{L\alpha}{2} \log \gamma_1$  in order for the sum rate constraint to be met with equality. It is also necessary to point out that if there exists  $i$  such that  $\gamma_1 > \gamma_i > 1$ , then  $R_T(\nu)$  in (32) is a function with more than one discontinuity point. Hence, it is possible that the sum rate constraint can not be met with equality for some  $R$  even if it is larger than the threshold  $\frac{L\alpha}{2} \log \gamma_1$ .

### B. Proof of Theorem V.2

The proof is constructed by contradiction: we assume that in the optimal rate allocation there exists an  $l$  such that

$$R_l^* = R_l < 1 \quad (37a)$$

$$R_{l+1}^* = R_{l+1} > 0, \quad (37b)$$

and show that such an assignment cannot be optimal. In particular, we show that there exists an assignment

$$R_l = \tilde{R}_l = \hat{R}_l + \gamma \quad (38a)$$

$$R_{l+1} = \tilde{R}_{l+1} = \hat{R}_{l+1} - \gamma \quad (38b)$$

for some small  $\gamma$  which provides a lower overall distortion.

From (15) and (16) we have that  $\mathbb{P}[\log F > 0] = \mathbb{P}[Q + K > 0]$  for

$$Q = \sum_{j=1, j \notin \{l, l+1\}}^L c_j (2U_j - 1), \quad (39)$$

and

$$K = c_l (2U_l - 1) + c_{l+1} (2U_{l+1} - 1). \quad (40)$$

We next show that there exists a choice of  $\delta_l, \delta_{l+1}$  in (38) such that using the same  $X$  estimate as if  $K$  was drawn according to (37) but for the rate assignment in (38) produces a lower distortion. In other words, we choose a new rate assignment, use a sub-optimal estimator but still obtain a lower distortion, from which it follows that the assignment in (37) cannot be optimal.

Consider first the case  $D_l > D_{l+1}$  and  $q_{l+1} > q_l$ , which is the case when the worse channel is described with a lower distortion than the better channel. In this case the distribution of  $K$  is

$$K = \begin{cases} -c_l - c_{l+1} & \bar{q}_l \bar{q}_{l+1} \\ -c_l + c_{l+1} & \bar{q}_l q_{l+1} \\ +c_l - c_{l+1} & q_l \bar{q}_{l+1} \\ +c_l + c_{l+1} & q_l q_{l+1} \end{cases} \quad (41)$$

with  $-c_l + c_{l+1} \leq 0$  since  $q_{l+1} > q_l$  implies  $c_l > c_{l+1}$ .

The assignment in (38) attains the distortions

$$\tilde{D}_l = \hat{D}_l + \delta_l \leq \hat{D}_l + \delta_{l+1} \quad (42a)$$

$$\tilde{D}_{l+1} = \hat{D}_{l+1} - \delta_{l+1} \quad (42b)$$

for some  $\delta_l, \delta_{l+1} > 0$  and where the inequality in (42a) follows since  $D_l > D_{l+1}$  and an increase in rate increases  $\tilde{D}_l$  more than a decrease in rate reduces  $D_{l+1}$ .

Next we want to show that there exists a choice of  $\gamma$  corresponding to a  $\delta_{l+1}$  for which the probability that  $F$  takes the positive values  $+c_l - c_{l+1}$  and  $+c_l + c_{l+1}$  is smaller with the assignment in (38) than with the assignment in (37). If we can find such a  $\gamma$ , then the assignment is certainly not optimal, since a suboptimal source estimator can produce a lower distortion. Given the symmetry in the distribution of  $K$ , this implies determining  $\delta_{l+1}$  so that  $q_l \bar{q}_{l+1} < \tilde{q}_l \tilde{q}_{l+1}$  and  $q_l q_{l+1} \leq \tilde{q}_l \tilde{q}_{l+1}$ . With the assignment in (42), we have that  $q_l \bar{q}_{l+1} < \tilde{q}_l \tilde{q}_{l+1}$ , so that the desired result is shown by showing that there exists a  $\gamma$  in (38) for which also  $q_l q_{l+1} \leq \tilde{q}_l \tilde{q}_{l+1}$ . This can be done by showing that the derivative of  $(p_l \star (D_l + \delta_{l+1})) (p_{l+1} \star (D_{l+1} - \delta_{l+1}))$  is decreasing in  $\delta_{l+1}$ . Indeed this derivative is equal to

$$\begin{aligned} & \frac{\partial (p_l \star (D_l + \delta_{l+1})) (p_{l+1} \star (D_{l+1} - \delta_{l+1}))}{\partial \delta_{l+1}} \\ &= q_l - q_{l+1} - 2(1 - 2p_l)(1 - 2p_{l+1})\delta_{l+1} - 2p_{l+1}(1 - 2p_l)D_l + 2p_l(1 - 2p_{l+1})D_{l+1}, \end{aligned} \quad (43)$$

which is negative when  $q_{l+1} > q_l$ ,  $p_{l+1} > p_l$  and  $D_l > D_{l+1}$  as per assumption.

Consider now the case in which  $q_{l+1} > q_l$  but  $D_l < D_{l+1}$ : also for this case the better channel is described with a lower distortion but an even lower distortion can be achieved by further reducing the distortion of the better channel. That is

$$\tilde{D}_l = \hat{D}_l + \delta_l \quad (44a)$$

$$\tilde{D}_{l+1} = \hat{D}_{l+1} - \delta_{l+1} \leq \hat{D}_{l+1} - \delta_l. \quad (44b)$$

Next we want to show that there exists a  $\delta_l$  for which  $(p_l \star (D_l + \delta_l)) (p_{l+1} \star (D_{l+1} - \delta_l))$  is decreasing in  $\delta_l$ . In this case the derivative becomes

$$\begin{aligned} & \frac{\partial (p_l \star (D_l + \delta_l)) (p_{l+1} \star (D_{l+1} - \delta_l))}{\partial \delta_l} \\ &= q_l - q_{l+1} + 2p_l(p_l - p_{l+1}) - 2(1 - 2p_l)(1 - 2p_{l+1})\delta_{l+1} - 2(p_{l+1} - p_l)(1 - 2p_l)D_l, \end{aligned} \quad (45)$$

which is again negative defined when  $q_{l+1} > q_l$  and  $p_l > p_{l+1}$  as assumed.

In the derivations above, we have shown that, regardless of the assignment in (37), one would obtain the choice policy as stated in the main theorem.